Numerical Optimization

Unconstrained Optimization

Shirish Shevade

Computer Science and Automation
Indian Institute of Science
Bangalore 560 012, India.

NPTEL Course on Numerical Optimization
Coordinate Descent Method

Consider the problem,

$$\min_{\mathbf{x}} f(\mathbf{x})$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $f \in C^1$.

Idea:

1. For every coordinate variable $x_i, i = 1, \ldots, n$, minimize $f(\mathbf{x})$ w.r.t. $x_i$, keeping the other coordinate variables $x_j, j \neq i$ constant.

2. Repeat the procedure in step 1 until some stopping condition is satisfied.
Coordinate Descent Method

(1) Initialize \( x^0, \epsilon \), set \( k := 0 \).

(2) while \( \|g^k\| > \epsilon \)

   for \( i = 1, \ldots, n \)
   
   \[ x_i^{\text{new}} = \arg\min_{x_i} f(x) \]
   
   \[ x_i = x_i^{\text{new}} \]

   endfor

endwhile

Output: \( x^* = x^k \), a stationary point of \( f(x) \).

- Globally convergent method if a search along any coordinate direction yields a unique minimum point.
Example: Consider the problem,

\[
\min_x f(x) \triangleq 4x_1^2 + x_2^2
\]

We use coordinate descent method with *exact line search* to solve this problem.

- \( x^0 = (-1, -1)^T \)
- Let \( d^0 = (1, 0)^T \)
- \( x^1 = x^0 + \alpha^0 d^0 \) where

\[
\alpha^0 = \arg \min_\alpha \phi_0(\alpha) \triangleq f(x^0 + \alpha d^0)
\]

- \( \phi_0(\alpha) = f \left( \begin{pmatrix} x_1^0 + \alpha d_1^0 \\ x_2^0 + \alpha d_2^0 \end{pmatrix} \right) = 4(\alpha - 1)^2 + 1 \)
- \( \phi'_0(\alpha) = 0 \Rightarrow \alpha^0 = 1 \Rightarrow x^1 = (0, -1)^T \)
- \( d^1 = (0, 1)^T, \ x^2 = x^1 + \alpha^1 d^1, \ \alpha^1 = \arg \min_\alpha \phi_1(\alpha) \triangleq \)

\[
f \left( \begin{pmatrix} 0 \\ \alpha - 1 \end{pmatrix} \right) = (\alpha - 1)^2 \Rightarrow \alpha^1 = 1 \Rightarrow x^2 = (0, 0)^T = x^* \]
\[ \min_x f(x) \triangleq 4x_1^2 + x_2^2 \]

For the above problem,

- Moving along coordinate directions and using exact lines search gives the solution in at most two steps.
- Same result is obtained even if \( d^0 \) and \( d^1 \) are interchanged.
Example: Consider the problem,

$$\min_{\mathbf{x}} f(\mathbf{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$

We use coordinate descent method with *exact line search* to solve this problem.

- $\mathbf{x}^0 = (-1, -1)^T$
- Let $\mathbf{d}^0 = (1, 0)^T$
- $\mathbf{x}^1 = \mathbf{x}^0 + \alpha^0 \mathbf{d}^0$ where

$$\alpha^0 = \arg \min_{\alpha} \phi_0(\alpha) \triangleq f(\mathbf{x}^0 + \alpha \mathbf{d}^0)$$

$$\phi_0(\alpha) = f \left( \begin{pmatrix} x_1^0 + \alpha d_1^0 \\ x_2^0 + \alpha d_2^0 \end{pmatrix} \right) = 4(\alpha - 1)^2 + 1 + 2(\alpha - 1)$$

$$\phi'_0(\alpha) = 0 \Rightarrow \alpha^0 = \frac{3}{4} \Rightarrow \mathbf{x}^1 = \left( -\frac{1}{4}, -1 \right)^T$$

- $\mathbf{d}^1 = (0, 1)^T$, $\mathbf{x}^2 = \mathbf{x}^1 + \alpha^1 \mathbf{d}^1$, $\alpha^1 = \arg \min_{\alpha} \phi_1(\alpha) \triangleq f \left( \begin{pmatrix} x_1^0 - \frac{1}{4} \\ 0 \end{pmatrix} \right) = (\alpha - 1)^2 + \frac{\alpha - 1}{2} + \frac{1}{4} \Rightarrow \alpha^1 = \frac{3}{4} \Rightarrow \mathbf{x}^2 = \left( -\frac{1}{4}, -\frac{1}{4} \right)^T \neq \mathbf{x}^*$
Example 1:
\[
\min_{\mathbf{x}} \ f_1(\mathbf{x}) \triangleq 4x_1^2 + x_2^2
\]
- \( \mathbf{H} = \begin{pmatrix} 8 & 0 \\ 0 & 2 \end{pmatrix} \).
- \( \mathbf{x}^* \), attained in \textit{at most two steps} using coordinate descent method

Example 2:
\[
\min_{\mathbf{x}} \ f_2(\mathbf{x}) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2
\]
- \( \mathbf{H} = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix} \).
- \( \mathbf{x}^* \), could not be attained in two steps using coordinate descent method (if \( \mathbf{x}^0 \) is not on one of the principal axes of the elliptical contours)
Consider the problem:

\[
\min_{x} f(x) \triangleq \frac{1}{2} x^T H x + c^T x
\]

where \(H\) is a symmetric positive definite matrix.

- Let \(\{d^0, d^1, \ldots, d^{n-1}\}\) be a set of linearly independent directions and \(x^0 \in \mathbb{R}^n\)
- Any \(x \in \mathbb{R}^n\) can be represented as

\[
x = x^0 + \sum_{i=0}^{n-1} \alpha^i d^i
\]

- Given \(\{d^0, d^1, \ldots, d^{n-1}\}\) and \(x^0 \in \mathbb{R}^n\), the given problem is to minimize \(\Psi(\alpha)\) defined as,

\[
\frac{1}{2} \left( x^0 + \sum_{i=0}^{n-1} \alpha^i d^i \right)^T H \left( x^0 + \sum_{i=0}^{n-1} \alpha^i d^i \right) + c^T \left( x^0 + \sum_{i=0}^{n-1} \alpha^i d^i \right)
\]
Define $D = (d^0 | d^1 | \ldots | d^{n-1})$ and $\alpha = (\alpha^0, \alpha^1, \ldots, \alpha^{n-1})$.

$$
\Psi(\alpha) = \frac{1}{2} \alpha^T D^T H D \alpha + (Hx^0 + c)^T D \alpha + \frac{1}{2} x^0^T H x^0 + c^T x^0
$$

$$
Q = D^T H D = \begin{pmatrix}
\quad d^0^T H d^0 & d^0^T H d^1 & \ldots & d^0^T H d^{n-1} \\
\quad d^1^T H d^0 & d^1^T H d^1 & \ldots & d^1^T H d^{n-1} \\
\quad \vdots & \vdots & \vdots & \vdots \\
\quad d^{n-1}^T H d^0 & d^{n-1}^T H d^1 & \ldots & d^{n-1}^T H d^{n-1}
\end{pmatrix}
$$

$Q$ will be diagonal matrix if $d^i^T H d^j = 0$, $\forall \ i \neq j$. 

Shirish Shevade | Numerical Optimization
Let $d^i T H d^j = 0$, $\forall i \neq j$.

$$Q = D^T H D = \begin{pmatrix}
    d^0 T H d^0 & 0 & \ldots & 0 \\
    0 & d^1 T H d^1 & \ldots & 0 \\
    \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \ldots & d^{n-1} T H d^{n-1}
\end{pmatrix}$$

Therefore,

$$Q_{ij}^{-1} = \begin{cases}
    \frac{1}{d^i T H d^j} & \text{if } j = i \\
    0 & \text{otherwise}
\end{cases}$$

$$\Psi(\alpha) = \frac{1}{2} (x^0 + \sum_i \alpha_i d^i)^T H (x^0 + \sum_i \alpha_i d^i) + c^T (x^0 + \sum_i \alpha_i d^i)$$

$$= \frac{1}{2} \sum_i \left[ (x^0 + \alpha_i d^i)^T H (x^0 + \alpha_i d^i) + 2c^T (x^0 + \alpha_i d^i) \right] + \text{constant}$$

- $\Psi(\alpha)$ is separable in terms of $\alpha^0, \alpha^1, \ldots, \alpha^{n-1}$
\[
\Psi(\alpha) = \frac{1}{2} \sum_i \left[ (x^0 + \alpha^i d^i)^T H (x^0 + \alpha^i d^i) + 2c^T (x^0 + \alpha^i d^i) \right]
\]

\[
\frac{\partial \Psi}{\partial \alpha^i} = 0 \Rightarrow \alpha^i = -\frac{d^i (Hx^0 + c)}{d^i H d^i}
\]

Therefore,
\[
x^* = x^0 + \sum_{i=0}^{n-1} \alpha^i d^i
\]

**Definition**

Let \( H \in \mathbb{R}^{n \times n} \) be a symmetric matrix. The vectors \( \{d^0, d^1, \ldots, d^{n-1}\} \) are said to be \( H \)-conjugate if they are linearly independent and \( d^i^T H d^j = 0 \ \forall \ i \neq j \).
Example: Consider the problem,

\[ \min_x f(x) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2 \]

- \( H = \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix} \)
- \( x^0 = (-1, -1)^T \)
- Let \( d^0 = (1, 0)^T \)
- \( x^1 = x^0 + \alpha^0 d^0 \) where

\[ \alpha^0 = \arg \min \alpha \phi_0(\alpha) \triangleq f(x^0 + \alpha d^0) \]

- \( \phi_0(\alpha) = f \left( \begin{pmatrix} x_1^0 + \alpha d_1^0 \\ x_2^0 + \alpha d_2^0 \end{pmatrix} \right) = 4(\alpha - 1)^2 + 1 + 2(\alpha - 1) \)
- \( \phi'_0(\alpha) = 0 \Rightarrow \alpha^0 = \frac{3}{4} \Rightarrow x^1 = (-\frac{1}{4}, -1)^T \)
- Choose a non-zero direction \( d^1 \) such that \( d^1^T H d^0 = 0 \)
- Let \( d^1 = (a, b)^T \). Therefore,

\[
\begin{pmatrix} a & b \end{pmatrix} \begin{pmatrix} 8 & -2 \\ -2 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = 0 \Rightarrow 8a - 2b = 0
\]
Let \( d^1 = (1, 4)^T \),

\[ x^2 = x^1 + \alpha^1 d^1 \]

where

\[ \alpha^1 = \arg\min_{\alpha} \phi_1(\alpha) \triangleq f \left( \frac{\alpha - \frac{1}{4}}{4\alpha - 1} \right) = \frac{3}{4} (4\alpha - 1)^2 \]

\[ \phi'_1(\alpha) = 0 \implies \alpha^1 = \frac{1}{4} \]

\[ x^2 = x^1 + \alpha^1 d^1 = (0, 0)^T = x^* \]

A convex quadratic function can be minimized in, \textit{at most}, \( n \) steps, provided we search along conjugate directions of the Hessian matrix.

Given \( H \), does a set of \( H \)-conjugate vectors exist? If yes, how to get a set of such vectors?
Conjugate Directions

Let $H \in \mathbb{R}^{n \times n}$ be a symmetric matrix.

- Do there exist $n$ conjugate directions w.r.t $H$?
  - $H$ is symmetric $\Rightarrow H$ has $n$ mutually orthogonal eigenvectors.

Let $v_1$ and $v_2$ be two orthogonal eigenvectors of $H$.
∴ $v_1^T v_2 = 0$.

$Hv_1 = \lambda_1 v_1 \Rightarrow v_2^T Hv_1 = \lambda_1 v_2^T v_1$
$\Rightarrow v_2^T Hv_1 = 0$
$\Rightarrow v_1$ and $v_2$ are $H$-conjugate

∴ $n$ orthogonal eigenvectors of $H$ are $H$-conjugate.
Conjugate Directions

- Let $H$ be a symmetric positive definite matrix and $d^0, d^1, \ldots, d^{n-1}$ be nonzero directions such that
  
  $$d^i^T Hd^j = 0, \ i \neq j.$$ 

  Are $d^0, d^1, \ldots, d^{n-1}$ linearly independent?

  $$\sum_{i=0}^{n-1} \mu^i d^i = 0 \Rightarrow \sum_{i=0}^{n-1} \mu^i d^T d^i = 0 \text{ for every } j = 0, \ldots, n - 1$$

  $$\Rightarrow \mu^j d^j^T Hd^j = 0$$

  $$\Rightarrow \mu^j = 0 \text{ for every } j = 0, \ldots, n - 1$$

  $$\Rightarrow d^0, d^1, \ldots, d^{n-1} \text{ are linearly independent}$$
Conjugate Directions

Geometric Interpretation:
Consider the problem:

$$\min_{x \in \mathbb{R}^2} \frac{1}{2} x^T H x + c^T x, \quad H \text{ symmetric positive definite matrix.}$$

Let $x^*$ be the solution. \( \therefore Hx^* = -c. \)
Let $x^0$ be any initial point. \( g^0 = Hx^0 + c \)
Let $d^0$ be some direction \( (d^0 \neq 0) \).
$x^1$ is found by doing exact line search along $d^0$. \( \therefore g^{1T}d^0 = 0. \)
$g^1 = Hx^1 + c.$

$$
(x^* - x^1)^T H d^0 = (Hx^* - Hx^1)^T d^0 \\
= -g^{1T} d^0 \\
= 0
$$

Therefore, the direction \( (x^* - x^1) \) is $H$ conjugate to $d^0$. 

Shirish Shevade  Numerical Optimization
Consider the problem:

\[
\min_x f(x) \triangleq \frac{1}{2} x^T H x + c^T x, \quad H \text{ symmetric positive definite matrix.}
\]

Let \( d^0, d^1, \ldots, d^{n-1} \) be \( H \)-conjugate. \( \therefore d^0, d^1, \ldots, d^{n-1} \) are linearly independent.

Let \( \mathcal{B}^k \) denote the subspace spanned by \( d^0, d^1, \ldots, d^{k-1} \).

Clearly, \( \mathcal{B}^k \subset \mathcal{B}^{k+1} \).

Let \( x^0 \in \mathbb{R}^n \) be any arbitrary point.

Let \( x^{k+1} = x^k + \alpha^k d^k \) where \( \alpha^k \) is obtained by doing exact line search:

\[
\alpha^k = \arg\min_{\alpha} f(x^k + \alpha d^k)
\]

Claim:

\[
x^k = \arg\min_x f(x) \quad \text{s.t.} \quad x \in x^0 + \mathcal{B}^k
\]
Exact line search:

\[ \alpha^k = \arg \min_{\alpha \in \mathbb{R}} f(x^k + \alpha d^k) \]

Therefore,

\[ \nabla f(x^k + \alpha^k d^k)^T d^k = 0 \Rightarrow g^{k+1T} d^k = 0 \quad \forall \ k = 0, \ldots, n - 1 \]

\[ x^k = x^{k-1} + \alpha^{k-1} d^{k-1} = x^j + \sum_{i=j}^{k-1} \alpha^i d^i \quad (j = 0, \ldots, k - 1) \]

\[ \therefore Hx^k + c = Hx^j + c + \sum_{i=j}^{k-1} \alpha^i Hd^i \]

\[ \therefore g^k = g^j + \sum_{i=j}^{k-1} \alpha^i Hd^i \]

\[ \therefore g^{kT} d^{j-1} = g^{jT} d^{j-1} + \sum_{i=j}^{k-1} \alpha^i d^i^T Hd^j = 0 \]

Therefore, \( g^{kT} d^j = 0 \quad \forall \ j = 0, \ldots, k - 1 \) or \( g^k \perp B^k \)
Note that for every \( j = 0, \ldots, n - 1 \),

\[
\alpha^j = \arg \min_{\alpha} \ f(x^j + \alpha d^j)
\]

\[
\therefore f(x^j + \alpha^j d^j) \leq f(x^j + \mu^j d^j), \quad \mu^j \in \mathbb{R}
\]

\[
\therefore f(x^j) + \alpha^j g^j d^j + \frac{1}{2} \alpha^j d^j T H d^j \leq f(x^j) + \mu^j g^j d^j + \frac{1}{2} \mu^j d^j T H d^j
\]

We need to show that \( f(x^k) \leq f(x) \) \( \forall x \in x^0 + B^k \) or

\[
f(x^0 + \sum_{j=0}^{k-1} \alpha^j d^j) \leq f(x^0 + \sum_{j=0}^{k-1} \mu^j d^j), \quad \mu^j \in \mathbb{R} \ \forall \ j.
\]

That is,

\[
f(x^0) + \sum_{j=0}^{k-1} (\alpha^j g^0 T d^j + \frac{1}{2} \alpha^j d^j T H d^j) \leq f(x^0) + \sum_{j=0}^{k-1} (\mu^j g^0 T d^j + \frac{1}{2} \mu^j d^j T H d^j)
\]

where \( \mu^j \in \mathbb{R} \ \forall \ j \).
For every $j = 0, \ldots, n - 1$,

$$f(x^j) + \alpha^j g^j d^j + \frac{1}{2} \alpha^j d^j T H d^j \leq f(x^j) + \mu^j g^j d^j + \frac{1}{2} \mu^j d^j T H d^j$$

Suppose $g^j d^j = g^0 d^j \forall j$

$$\therefore \alpha^j g^0 d^j + \frac{1}{2} \alpha^j d^j T H d^j \leq \mu^j g^0 d^j + \frac{1}{2} \mu^j d^j T H d^j \quad \forall j$$

Therefore,

$$f(x^0) + \sum_{j=0}^{k-1} (\alpha^j g^0 d^j + \frac{1}{2} \alpha^j d^j T H d^j) \leq f(x^0) + \sum_{j=0}^{k-1} (\mu^j g^0 d^j + \frac{1}{2} \mu^j d^j T H d^j)$$

$$\therefore f(x^0 + \sum_{j=0}^{k-1} \alpha^j d^j) \leq f(x^0 + \sum_{j=0}^{k-1} \mu^j d^j), \quad \mu^j \in \mathbb{R} \forall j$$

$$\therefore f(x^k) \leq f(x), \quad \forall x \in x^0 + B^k$$
We need to show that

\[ g_j^T d_j = g_0^T d_j \quad \forall j \]

Consider, \( x^j = x^0 + \sum_{i=0}^{j-1} \alpha^i d^i \).

\[
\therefore Hx^j + c = Hx^0 + c + \sum_{i=0}^{j-1} \alpha^i Hd^i
\]

\[
\therefore g^j = g^0 + \sum_{i=0}^{j-1} \alpha^i Hd^i
\]

\[
\therefore g_j^T d_j = g_0^T d_j + \sum_{i=0}^{j-1} \alpha^i d^iT Hd^j
\]

\[
\therefore g_j^T d_j = g_0^T d_j \quad \forall j
\]
Expanding Subspace Theorem

Consider the problem to minimize $f(x) \triangleq \frac{1}{2} x^T H x + c^T x$ where $H$ is symmetric positive definite matrix. Let $d^0, d^1, \ldots, d^{n-1}$ be $H$-conjugate and let $x^0 \in \mathbb{R}^n$ be any initial point. Let

$$\alpha^k = \arg \min_{\alpha \in \mathbb{R}} f(x^k + \alpha d^k), \forall k = 0, \ldots, n - 1$$

and

$$x^{k+1} = x^k + \alpha^k d^k, \forall k = 0, \ldots, n - 1.$$

Then, for all $k = 0, \ldots, n - 1$,

1. $g^T d^j = 0, j = 0, \ldots, k$
2. $g^T d^k = g^0 d^k$
3. $x^{k+1} = \arg \min_x f(x)$

s.t. $x \in x^0 + B^k$
Given a set of $n$ directions, $d^0, d^1, \ldots, d^{n-1}$ which are $H$-conjugate and $x^0 \in \mathbb{R}^n$, it is easy to determine $\alpha^{i*}$, $\forall \ i = 0, \ldots, n - 1$,

$$
\alpha^{i*} = -\frac{d^iT(Hx^0 + c)}{d^iT Hd^i}
$$

and get

$$
x^* = x^0 + \sum_{i=0}^{n-1} \alpha^{i*} d^i
$$

- How do we construct the $H$-conjugate directions, $d^0, d^1, \ldots, d^{n-1}$?
- Given the $H$-conjugate directions, $d^0, d^1, \ldots, d^{k-1}$, how do we determine $\alpha^k$ where

$$
\alpha^k = \arg \min_{\alpha} f(x^k + \alpha d^k)
$$
\[ x^* - x^0 = \sum_{i=0}^{n-1} \alpha^i d^i \]

\[ \therefore d^T H(x^* - x^0) = \alpha^k d^T H d^k \]

\[ \therefore \alpha^k = \frac{d^T H(x^* - x^0)}{d^T H d^k} \]

Suppose that after \( k \) iterative steps and obtaining \( k H \)-conjugate directions,

\[ x^k - x^0 = \sum_{i=0}^{k-1} \alpha^i d^i \]

\[ \therefore d^T H(x^k - x^0) = 0 \]
Given, $d^k H(x^k - x^0) = 0$,

$$
\therefore \alpha^k = \frac{d^k H(x^* - x^k + x^k - x^0)}{d^k H d^k}
$$

$$
= \frac{d^k (Hx^* - Hx^k)}{d^k H d^k}
$$

$$
= \frac{d^k (-c - Hx^k)}{d^k H d^k}
$$

$$
= - \frac{g^k d^k}{d^k H d^k}
$$

Therefore,

$$
\alpha^k = - \frac{g^k d^k}{d^k H d^k}
$$
Suppose \( \{ -g^0, -g^1, \ldots, -g^{n-1} \} \) is a linearly independent set of vectors.

Use Gram-Schmidt procedure to determine the \( H \)-conjugate vectors, \( d^0, d^1, \ldots, d^{n-1} \).

- Let \( d^0 = -g^0 \)
- In general,

\[
d^k = -g^k + \sum_{j=0}^{k-1} \beta^j d^j, \quad k = 1, \ldots, n - 1
\]

But we want \( d^0, d^1, \ldots, d^{n-1} \) to be \( H \)-conjugate vectors.

\[
d^i^T H d^k = -d^i^T H g^k + \sum_{j=0}^{k-1} \beta^j d^i^T H d^j, \quad i = 0, \ldots, k - 1
\]

\[
\therefore 0 = -d^i^T H g^k + \beta^i d^i^T H d^i, \quad i = 0, \ldots, k - 1
\]

\[
\therefore \beta^i = \frac{g^k^T H d^i}{d^i^T H d^i}
\]
\[ \therefore \mathbf{d}^k = -\mathbf{g}^k + \sum_{j=0}^{k-1} \left( \frac{\mathbf{g}^k \mathbf{T} \mathbf{H} \mathbf{d}^j}{d^j \mathbf{d}^j \mathbf{T} \mathbf{H} \mathbf{d}^j} \right) \mathbf{d}^j \]

We now need to show that \( \{ -\mathbf{g}^0, -\mathbf{g}^1, \ldots, -\mathbf{g}^{n-1} \} \) is a \textit{linearly independent} set of vectors.

Note that
\[
\text{span}\{ \mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{k-1} \} = \text{span}\{ -\mathbf{g}^0, -\mathbf{g}^1, \ldots, -\mathbf{g}^{k-1} \}
\]

We have already shown that
\[
\{ \mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{k-1} \} \text{ are } \mathbf{H}\text{-conjugate} \Rightarrow \mathbf{g}^k \perp \mathcal{B}^k
\]
\[ \therefore -\mathbf{g}^k \perp \text{span}\{ \mathbf{d}^0, \mathbf{d}^1, \ldots, \mathbf{d}^{k-1} \} \]
\[ \therefore -\mathbf{g}^k \perp \text{span}\{ -\mathbf{g}^0, -\mathbf{g}^1, \ldots, -\mathbf{g}^{k-1} \} \]

Therefore, \( \{ -\mathbf{g}^0, -\mathbf{g}^1, \ldots, -\mathbf{g}^{n-1} \} \) is a \textit{linearly independent} set of vectors.
Now, consider
\[
\begin{align*}
    d^0 &= -g^0 \\
    d^k &= -g^k + \sum_{j=0}^{k-1} \left( \frac{g^k T H d^j}{d^j T H d^j} \right) d^j \quad \forall \ k = 1, \ldots, n - 1
\end{align*}
\]

Note that \( x^{j+1} = x^j + \alpha^j d^j \) and \( g^{j+1} = g^j + \alpha^j H d^j \).
Therefore,
\[
    H d^j = \frac{1}{\alpha^j} (g^{j+1} - g^j)
\]
Thus,
\[
\begin{align*}
    d^k &= -g^k + \sum_{j=0}^{k-1} \left( \frac{g^k T (g^{j+1} - g^j)}{d^j T (g^{j+1} - g^j)} \right) d^j \\
    &= -g^k + \left( \frac{g^k T g^k}{d^{k-1} T (g^k - g^{k-1})} \right) d^{k-1}
\end{align*}
\]
\[ d^k = -g^k + \left( \frac{g^k T g^k}{d^{k-1T} (g^k - g^{k-1})} \right) d^{k-1} \]

Due to exact line search, \( g^k T d^{k-1} = 0 \).

\[ d^{k-1} = -g^{k-1} + \beta^{k-2} d^{k-2} \]

\[ -d^{k-1T} g^{k-1} = g^{k-1T} g^{k-1} + \beta^{k-2} g^{k-1T} d^{k-2} \]

Therefore,

\[ d^k = -g^k + \frac{g^k T g^k}{g^{k-1T} g^{k-1}} d^{k-1}, \quad k = 1, \ldots, n - 1 \]

Fletcher-Reeves method
Conjugate Gradient Algorithm (Fletcher-Reeves)

For Quadratic function, $\frac{1}{2}x^THx + c^Tx$, $H$ symmetric positive definite

(1) Initialize $x^0, \epsilon, d^0 = -g^0$, set $k := 0$.

(2) while $\|g^k\| > \epsilon$

(a) $\alpha^k = -\frac{g_k^T d_k}{d_k^T H d_k}$

(b) $x^{k+1} = x^k + \alpha^k d^k$

(c) $g^{k+1} = Hx^{k+1} + c$

(d) $\beta^k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}$

(e) $d^{k+1} = -g^{k+1} + \beta^k d^k$

(f) $k := k + 1$

endwhile

Output: $x^* = x^k$, global minimum of $f(x)$.  

Shirish Shevade  |  Numerical Optimization
Example:

$$\min f(x) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$

Conjugate Gradient algorithm (Fletcher-Reeves) with exact line search applied to $f(x)$
Example:

$$\min f(x) \triangleq 4x_1^2 + x_2^2 - 2x_1x_2$$

Conjugate Gradient algorithm (Fletcher-Reeves) with exact line search applied to $f(x)$
Extension to Nonquadratic function, $f(x)$:

**Conjugate Gradient Algorithm (Fletcher-Reeves)**

1. Initialize $x^0$, $\epsilon$, $d^0 = -g^0$, set $k := 0$.
2. while $\|g^k\| > \epsilon$
   
   a. $\alpha^k = \arg\min_{\alpha > 0} f(x^k + \alpha d^k)$
   
   b. $x^{k+1} = x^k + \alpha^k d^k$
   
   c. Compute $g^{k+1}$
   
   d. if $k < n - 1$
      
      i. $\beta^k = \frac{g^{k+1T}g^{k+1}}{g^{kT}g^k}$
      
      ii. $d^{k+1} = -g^{k+1} + \beta^k d^k$
      
      iii. $k := k + 1$
   
   else
      
      i. $x^0 = x^{k+1}$
      
      ii. $d^0 = -g^{k+1}$
      
      iii. $k := 0$
   
   endif

endwhile

Output: $x^* = x^k$, a stationary point of $f(x)$. 

Shirish Shevade Numerical Optimization
\[ \beta^k \text{ Determination} \]

- **Fletcher-Reeves method**
  \[ \beta_{FR}^k = \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \]

- **Polak-Ribiere method**
  \[ \beta_{PR}^k = \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T g_{k-1}} \]

- **Hestenes-Steifel method**
  \[ \beta_{HS}^k = \frac{g_k^T (g_k - g_{k-1})}{(g_k - g_{k-1})^T d_{k-1}} \]
\[
B_{BFGS}^k = B + \left( 1 + \frac{\gamma^T B \gamma}{\delta^T \gamma} \right) \frac{\delta \delta^T}{\delta^T \gamma} - \left( \frac{\delta \gamma^T B + B \gamma \delta^T}{\delta^T \gamma} \right)
\]

Memoryless BFGS iteration

\[
B_{BFGS}^k = I + \left( 1 + \frac{\gamma^T \gamma}{\delta^T \gamma} \right) \frac{\delta \delta^T}{\delta^T \gamma} - \left( \frac{\delta \gamma^T + \gamma \delta^T}{\delta^T \gamma} \right)
\]

With exact line search, \( \delta^{k-1} g^k = \alpha^{k-1} d^{k-1} g^k = 0 \). Therefore,

\[
d_{BFGS}^k = -B_{BFGS}^k g^k = -g^k + \frac{\delta \gamma^T g^k}{\delta^T \gamma} = -g^k + \frac{g^k (g^k - g^{k-1})}{(g^k - g^{k-1})^T d^{k-1}} d^{k-1}
\]

\[\beta_{HS}^k\]
For nonquadratic function, $f(x)$:

**Conjugate Gradient Algorithm (Fletcher-Reeves)**

(1) Initialize $x^0$, $\epsilon$, $d^0 = -g^0$, set $k := 0$.

(2) while $\|g^k\| > \epsilon$

   (a) $\alpha^k = \arg\min_{\alpha > 0} f(x^k + \alpha d^k)$

   (b) $x^{k+1} = x^k + \alpha^k d^k$

   (c) Compute $g^{k+1}$

   (d) if $k < n - 1$

      (i) $\beta^k = \frac{g^{k+1T}g^{k+1}}{g^kTg_k}$

      (ii) $d^{k+1} = -g^{k+1} + \beta^k d^k$

      (iii) $k := k + 1$

    else

      (i) $x^0 = x^{k+1}$

      (ii) $d^0 = -g^{k+1}$

      (iii) $k := 0$

    endif

endwhile

Output: $x^* = x^k$, a stationary point of $f(x)$. 

Shirish Shevade  
Numerical Optimization